CENTER FOR STOCHASTIC PROCESSES

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina

ON A BASIS FOR "PEAKS OVER THRESHOLD" MODELING

by

M.R. Leadbetter

Technical Report No. 285

March 1990

219. J. Bather, Stopping rules and ordered families of distributions, Dec. 87. *Sequential Anal.*, 7, 1988, 111-126.

220. S. Cambanis and M. Maejima, Two classes of self-similar stable processes with stationary increments, Jan. 88. *Stochastic Proc. Appl.*, 32, 1989, 305-329.

221. H.P. Hucke, G. Kallianpur and R.L. Karandikar, Smoothness properties of the conditional expectation in finitely additive white noise filtering, Jan. 88. *J. Multivariate Anal.*, 27, 1988, 261-269.

222. I. Mitoma, Weak solution of the Langevin equation on a generalized functional space, Feb. 88. (Revised as No. 238).

223. L. de Haan, S.I. Resnick, H. Rootzén and C. de Vries, Extremal behaviour of solutions to a stochastic difference equation with applications to arch-processes, Feb. 88. *Stochastic Proc. Appl.*, 32, 1989, 213-224.

224. O. Kallenberg and J. Szulga, Multiple integration with respect to Poisson and Lévy processes, Feb. 88. *Prob. Theor. Rel. Fields*, 83, 1989, 101-134.

225. D.A. Dawson and L.G. Gorostiza, Generalized solutions of a class of nuclear space valued stochastic evolution equations, Feb. 88. *Appl. Math. Optimization*, to appear.

226. G. Samorodnitsky and J. Szulga, An asymptotic evaluation of the tail of a multiple symmetric α-stable integral, Feb. 88. *Ann. Probability*, 17, 1989, 1503-1520.

227. J.J. Hunter, The computation of stationary distributions of Markov chains through perturbations, Mar. 88.

228. H.C. Ho and T.C. Sun, Limiting distribution of nonlinear vector functions of stationary Gaussian processes, Mar. 88. *Ann. Probability*, to appear.

229. R. Brigola, On functional estimates for ill-posed linear problems, Apr. 88.

230. M.R. Leadbetter and S. Nandagopalan, On exceedance point processes for stationary sequences under mild oscillation restrictions, Apr. 88. *Proc. Oberwolfach Conf. on Extremal Value Theory*, J. Hüsler and R. Reiss, eds., Springer, 1989, 69-80.

231. S. Cambanis, J. P. Nolan and J. Rosinski, On the oscillation of infinitely divisible processes, Apr. 88. *Stochastic Proc. Appl.*, 1990, to appear.

232. G. Hardy, G. Kallianpur and S. Ramasubramanian, A nuclear space-valued stochastic differential equation driven by Poisson random measures, Apr. 88.

233. D.J. Daley, T. Rolski, Light traffic approximations in queues (II), May 88. *Math. Operat. Res.*, to appear.

234. G. Kallianpur, I. Mitoma, R.L. Wolpert, Nuclear space-valued diffusion equations, July 88. *Stochastics*, 1989, to appear.

235. S. Cambanis, Admissible translates of stable processes: A survey and some new models, July 88.

236. E. Platen, On a wide range exclusion process in random medium with local jump intensity, Aug. 88.

237. R.L. Smith, A counterexample concerning the extremal index, Aug. 88. *Adv. Appl. Probab.*, 20, 1988, 681-683.

238. G. Kallianpur and I. Mitoma, A Langevin-type stochastic differential equation on a space of generalized functionals, Aug. 88.

239. C. Houdré, Harmonizability, V-boundedness, (2,P)-boundedness of stochastic processes, Aug. 88. *Prob. Th. Rel. Fields*, 84, 1990, 39-54.

240. G.W. Johnson and G. Kallianpur, Some remarks on Hu and Meyer's paper and infinite dimensional calculus on finitely additive canonical Hilbert space, Sept. 88. *Th. Prob. Appl.*, to appear.

241. L. de Haan, A Brownian bridge connected with extreme values, Sept. 88. *Sankhya*, 1989, to appear.

242. O. Kallenberg, Exchangeable random measures in the plane, Sept. 88, *J. Theor. Probab.*, 3, 1990, 81-136.

243. E. Masry and S. Cambanis, Trapezoidal Monte Carlo integration, Sept. 88. *SIAM J. Numer. Anal.*, 27, 1990, 225-246.

244. L. Pitt, On a problem of H.P. McKean, Sept. 88. *Ann. Probability*, 17, 1989, 1651-1657.

245. C. Houdré, On the linear prediction of multivariate ?)-bounded processes, Sept. 88.

246. C. Houdré, Stochastic processes as Fourier integrals and dilation of vector measures, Sept. 88, *Bull. Amer. Math. Soc.*, 21, 1989, 281-285.

247. J. Mijnheer, On the rate of convergence in Strassen's functional law of the iterated logarithm, Sept. 88, *Probab. Theor. Rel. Fields*, to appear.

248. G. Kallianpur and V. Perez-Abreu, Weak convergence of solutions of stochastic evolution equations on nuclear spaces, Oct. 88, Proc. Trento Conf. on Infinite Dimensional Stochastic Differential Equations, 1989, to appear.

249. R.L. Smith, Bias and variance approximations for estimators of extreme quantiles, Nov. 88.

250. H.L. Hurd, Spectral coherence of nonstationary and transient stochastic processes, Nov. 88. 4th Annual ASSP Workshop on Spectrum Estimation and Modeling, Minneapolis, 1988, 387-390.

251. J. Leskow, Maximum likelihood estimator for almost periodic stochastic processes models, Dec. 88.

252. M.R. Leadbetter and T. Hsing, Limit theorems for strongly mixing stationary random measures, Jan. 89. *Stochastic Proc. Appl.*, to appear.

253. M.R. Leadbetter, I. Weissman, L. de Haan, H. Rootzén, On clustering of high values in statistically stationary series, Jan. 89. *Proc. 4th Int. Meeting on Stat. Climatology*, 217-222.

254. J. Leskow, Least squares estimation in almost periodic point processes models, Feb. 89.

255. N.N. Vakhania, Orthogonal random vectors and the Hurwitz-Radon-Eckmann theorem, Apr. 89.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION
Unclassified

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY
N/A

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE
N/A

3. DISTRIBUTION / AVAILABILITY OF REPORT
Approved for Public Release;
Distribution Unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
Technical Report No. 285

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AFOSR-TR- 90 0651

6a. NAME OF PERFORMING ORGANIZATION
University of North Carolina
Center for Stochastic Processes

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
AFOSR/NM

6c. ADDRESS *(City, State, and ZIP Code)*
Statistics Department
CB #3260, Phillips Hall
Chapel Hill, NC 27599-3260

7b. ADDRESS *(City, State, and ZIP Code)*
Bldg. 410
Bolling Air Force Base, DC 20332-6448

8a. NAME OF FUNDING / SPONSORING ORGANIZATION
AFOSR

8b. OFFICE SYMBOL *(If applicable)*
NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
F49620 85C 0144

8c. ADDRESS *(City, State, and ZIP Code)*
Bldg. 410
Bolling AFB, DC 20332-6448

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO
6.1102F	2304	A5	

11. TITLE *(Include Security Classification)*
On a basis for "peaks over threshold" modeling

12. PERSONAL AUTHOR(S)
Leadbetter, M.R.

13a. TYPE OF REPORT
preprint

13b. TIME COVERED
FROM ________ TO ______

14. DATE OF REPORT *(Year, Month, Day)*
1990, March

15. PAGE COUNT
10

16. SUPPLEMENTARY NOTATION
N/A

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
XXXXX	XXXXXXXXX	XXX

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

(Please see over)

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
Professor Eytan Barouch

22b. TELEPHONE *(Include Area Code)*
(202)767-5026

22c. OFFICE SYMBOL
AFOSR/NM

DD Form 1473, JUN 86 *Previous editions are obsolete*

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFED

19. Abstract

Abstract. "Peaks over Threshold" ("POT") models commonly used e.g. in hydrology, assume that peak values of an iid or stationary sequence X_i above a high value u, occur at Poisson points, and the excess values of the peak above u are independent with an arbitrary common d.f. G. Motivation for these models has been provided by R.L. Smith (cf. [7],[8]), by using Pareto-type approximations of Pickands ([6]) for distributions of such excess values. These works strongly suggest that the Pareto family provides the appropriate class of distributions G for the POT model.

In the present paper we consider the point process of excess values of peaks above a high level u and demonstrate that this converges in distribution to a Compound Poisson Process as $u \longrightarrow \infty$ under appropriate assumptions. It is shown that the multiplicity distribution of this limit (i.e. the limiting distribution of excess values of peaks) must belong to the Pareto family and detailed forms are given for the normalizing constants involved. This exhibits the POT model specifically as a limit for the point process of excesses of peaks and delineates the distributions involved.

ON A BASIS FOR "PEAKS OVER THRESHOLD" MODELING

by

M.R. Leadbetter

University of North Carolina

(sub i

Abstract. "Peaks over Threshold" ~~("POT")~~ models commonly used e.g. in hydrology, assume that peak values of an iid or stationary sequence X_i above a high value u, occur at Poisson points, and the excess values of the peak above u are independent with an arbitrary common d.f. G. Motivation for these models has been provided by ~~R.L.~~ Smith ~~(cf. [7],[8])~~ by using Pareto-type approximations of Pickands ~~([6])~~ for distributions of such excess values. These works strongly suggest that the Pareto family provides the appropriate class of distributions G for the POT model.

for ~~In the present~~ paper ~~we~~ considers the point process of excess values of peaks above a high level u and demonstrate that this converges in distribution to a Compound Poisson Process as u approaches limit of infinity ~~$\to \infty$~~ under appropriate assumptions. It is shown that the multiplicity distribution of this limit (i.e. the limiting distribution of excess values of peaks) must belong to the Pareto family and detailed forms are given for the normalizing constants involved. This exhibits the POT model specifically as a limit for the point process of excesses of peaks and delineates the distributions involved.

(KR)

Research sponsored by the Air Force Office of Scientific Research Contract No. F49260 85C 0144.

1. Introduction.

In what are sometimes called "Peaks over Threshold" (POT) models (cf. [7]), the excess values over a high level u by an observed time series are assumed to occur at Poisson points and to have arbitrary common distributions. That is if $\{X_i: i=1,2,\ldots\}$ is e.g. a stationary (or iid) sequence, the exceedance points $\{i: X_i > u\}$ are assumed to be Poisson and the corresponding excess values $(X_i-u)_+$ to be independent with an arbitrary distribution.

The Poisson nature of the occurrence of exceedances is intuitively clear since e.g. if X_i are iid with d.f. F, the number of exceedances of a high level u_n by $X_1,\ldots,X_n$, is binomial with parameters $(n, 1-F(u_n))$ and hence approximately Poisson if $n(1-F(u_n))$ converges to some value $\tau > 0$. This will be made more precise below by a time normalization. Further motivation for the model is provided by R.L. Smith ([7],[8]) based on theory of Pickands [6], restricting the distribution for excess values to a "generalized Pareto" (GP) form

$$
\begin{aligned}
(1.1) \qquad G_{\alpha,\beta}(x) &= 1 - (1 + \alpha x/\beta)^{-1/\alpha} \qquad \beta > 0,\ \alpha \neq 0 \\
&= 1 - e^{-x/\beta} \qquad \beta > 0,\ \alpha = 0
\end{aligned}
$$

where the range of x is $(0,\infty)$ if $\alpha \geq 0$ and $(0,-\alpha^{-1}\beta)$ if $\alpha < 0$.

This class is a flexible 2-parameter family, but more importantly for a wide class of F's the excess distribution

$$(1.2) \qquad F_u(x) = P\{X - u \leq x | X > u\}$$

is approximately GP in a sense shown by Pickands [6], viz.

$$(1.3) \qquad \inf_{\beta} \sup_{x} |F_u(x) - G_{\alpha,\beta}(x)| \longrightarrow 0 \quad \text{as } u \longrightarrow \infty$$

for some fixed α. That is for high levels u, β may be chosen (depending on u) so that $G_{\alpha,\beta}(x)$ is uniformly close to $F_u(x)$.

Write x_F for the right endpoint (sup{x: F(x) < 1} of the d.f. F and $\overline{F}(x) = 1-F(x)$, the tail of F. Then the class of d.f.'s F for which the above GP approximation holds includes all those satisfying

$$(1.4) \qquad \overline{F}(u+xg(u))/\overline{F}(u) \longrightarrow \overline{G}(x) \quad \text{as } u \longrightarrow x_F$$

for some function $g(u) > 0$, some d.f. G and all $0 < x < x_F$ ($\leq \infty$). It is known ([6]) that any such G in (1.4) must be G.P. as in (1.1) for some α,β.

Note that (1.4) holds for all d.f.'s F of interest in extreme value theory i.e. such that if $M_n = \max(X_1,\dots,X_n)$, $P\{a_n(M_n-b_n) \leq x\}$ (= $F^n(x/a_n+b_n)$) has a non degenerate limit $\Lambda(x)$. For example if $\overline{G}(x) = e^{-x}$, (1.1) is a classical domain of attraction criterion for a "Type I" extreme value distribution $\Lambda(x) = \exp(-e^{-x})$. If instead F has a regulary varying tail $(1-F(ux))/(1-F(u)) \longrightarrow x^{-\alpha}$, $\alpha>0$, as $u \longrightarrow \infty$, each $x > 0$, then (1.1) holds with $\overline{G}(x) = (1+x)^{-\alpha}$, $g(u) = u$, and $\Lambda(x)$ is then "Type II" i.e. $\Lambda(x) = \exp(-x^{-\alpha})$, $x > 0$.

Hence in a wide variety of cases of interest the distribution $F_u(x)$ of excesses (given by (1.2)) of the level u is approximately GP, $G_{\alpha,\beta}(x)$ in the sense stated, where α is fixed but β can change with u. As discussed in [7] this provides significant intuitive support for the POT model. Our purpose here is to further justify the model by exhibiting it as the limit of point processes of excesses of high levels, the limiting distribution of excess values being shown to be GP, $G_{\alpha,\beta}(x)$ where now β as well as α, is independent of u.

For clarity this will be shown for iid sequences in Section 2 and extended to dependent (mixing) situations in Section 3. In the latter case high serial correlation can cause clustering of exceedances of high levels and the peak values are then defined to be the largest in each cluster. Two points worth noting are (i) the dependence modifies the theory via the introduction of a single parameter, the "extremal index" (essentially the inverse of mean cluster size) and (ii) the GP form applies to the peak values but not necessarily to other cluster properties such as their lengths. A corresponding theory will be indicated in Section 3 for other such cases.

2. The iid case.

Let $\{u_n\}$ be a sequence of levels such that

$$n(1-F(u_n)) \longrightarrow \tau > 0. \tag{2.1}$$

Define point processes N_n to consist of the points i/n for which $X_i > u_n$, i.e. the exceedance points normalized by $1/n$. N_n is thus defined on the positive real line but it will be convenient to restrict attention to the unit interval, corresponding to exceedances among the n sample values $X_1, \dots, X_n$. Hence $N_n(B)$ is defined for (Borel) subsets $B \subset (0,1]$ by $N_n(B) = \#\{i/n \in B: X_i > u_n\}$.

It is trivial to show that under (2.1) that $N_n \xrightarrow{d} N$ where N is a Poisson Process on (0,1] with intensity τ. For if $I = (a,b] \subset (0,1]$, $N_n(I)$ is binomial with parameters $[nb]-[na]$, $p_n = 1-F(u_n)$ ([ ] denoting integer part) and $np_n \longrightarrow \tau$ so that

$$P\{N_n(I) = r\} \longrightarrow e^{-\tau(b-a)}[\tau(b-a)]^r/r! = P\{N(I) = r\}.$$

Hence $N_n(I) \xrightarrow{d} N(I)$ and by independence if $I_1, \dots, I_k$ are disjoint

$$(N_n(I_1), \dots N_n(I_k)) \xrightarrow{d} (N(I_1) \dots N(I_k))$$

which is sufficient ([2]) to show full weak convergence $N_n \xrightarrow{d} N$.

Now associate with each point of N_n (i/n such that $X_i > u_n$) the corresponding excess value $X_i - u_n$ to give the "point process N_n^* of excesses".

Technically perhaps N_n^* should be regarded as a "marked point process" (or as an atomic random measure) since the (X_i-u_n) are not necessarily integer valued but it will be convenient (and legitimate) to call it a point process whose events (at $(i/n: X_i > u_n)$) have (not necessarily integer) multiplicities (X_i-u_n).

Since, as above, the positions of the excesses converge to a Poisson Process, it seems evident that one should expect any limit N^* for N_n^* to be a compound Poisson process having events at Poisson points with intensity τ and (independent) multiplicities with some common d.f. G. Such a result may be shown (cf. [1]) but here we give sufficient conditions for such convergence. Here and below we write $N^* = CP(\tau,G)$ to denote a Compound Poisson point process whose Poisson events have intensity τ with (not necessarily integer-valued) multiplicities having d.f. G.

<u>Theorem 2.1</u>. Let X_i, i=1,2... be iid with d.f. F satisfying (1.4) for some g, G, and let (u_n) be levels satisfying (2.1). Then $a_nN_n^* \xrightarrow{d} N^*$, $CP(\tau,G)$ where τ is as in (2.1), G as in (1.4) and $a_n = 1/g(u_n)$.

<u>Proof</u>: If $G_n(x)$ $(= F_{u_n}(x/a_n))$ denotes the conditional d.f. of $a_n(X_1-u_n)$ given $X_1 > u_n$, then clearly

$$\mathcal{E}\exp\{-sa_n(X_1-u_n)_+\} = F(u_n) + \bar{F}(u_n) \int_0^\infty e^{-sx}dG_n(x).$$

But from (1.4), $G_n(x) \longrightarrow G(x)$ and hence $\int_0^\infty e^{-sx}dG_n(x) \longrightarrow \phi(s) = \int_0^\infty e^{-sx}dG(x)$ so that if $I = (a,b] \subset (0,1]$ contains m_n points i/n $(m_n \sim n(b-a))$

$$\mathcal{E}^{m_n}\exp\{-sa_n(X_1-u_n)_+\} = [1-(\tau/n)(1-\phi(s))(1+o(1))]^{m_n}$$

$$\longrightarrow e^{-\tau(b-a)(1-\phi(s))}$$

which is the Laplace Transform $\mathcal{E}e^{-sN^*(a,b]}$ where N^* is $CP(\tau,G)$. Since $N_n^*(I)$ is the sum $\sum_{i/n \in I} (X_i - u_n)_+$ of m_n iid terms, it follows that $a_n N_n^*(I) \xrightarrow{d} N^*(I)$. Hence by independence

$$(a_n N_n^*(I_1) \ldots a_n N_n^*(I_n)) \xrightarrow{d} (N_n(I_1) \ldots N_n(I_k))$$

for any disjoint $I_1, \ldots, I_k$ so that $a_n N_n^* \xrightarrow{d} N^*$ as required. □

Thus in the iid case the (normalized) point process $a_n N_n^*$ of excesses over u_n has a Compound Poisson limit and may be regarded as approximately $CP(\tau, G)$ for large n. Note again that G is a GP distribution $G_{\alpha,\beta}$ for some fixed α, β. This provides a strong basis for the POT model in the iid case.

3. Stationary sequences

A stationary sequence $X_1, X_2, \ldots$ can exhibit both "long range" and "local" dependence between the X_i. Here the former, but not the latter, will be restricted by a "mixing condition" which enables the collection of X_i's into groups which are approximately independent but with possible high dependence within groups. Strong mixing will suffice to restrict long range dependence. However one may "tailor" this to the problem at hand with a slightly weaker restriction $\Delta(v_n)$ defined for any sequence of constants v_n as follows. For $1 \le j \le k \le n$ write $\mathcal{B}_{jk}(v_n) = \sigma\{(X_s - v_n)_+ : j \le s \le k\}$ and say that $\Delta(v_n)$ holds if for $1 \le \ell < n$ $|P(A \cap B) - P(A)P(B)| \le \alpha_{n,\ell}$ whenever $A \in \mathcal{B}_{1,j}(v_n)$, $B \in \mathcal{B}_{j+\ell,n}(v_n)$, $1 \le j < n-\ell$ and $\alpha_{n,\ell_n} \longrightarrow 0$ for some $\ell_n = o(n)$.

High <u>local</u> dependence is reflected in the presence of clustering of exceedances of high levels. "Clusters" will be defined precisely below, but we first note that for levels (u_n) satisfying (2.1) the mean size of a cluster typically converges to a parameter whose inverse θ is sometimes called the "extremal index" of the sequence $\{X_i\}$. Specifically θ is defined by the property that if $M_n = \max(X_1, X_2, \ldots, X_n)$ and u_n satisfies (2.1) then

$$P(M_n \leq u_n) \longrightarrow e^{-\theta\tau}, \tag{3.1}$$

θ being independent of τ. (For i.i.d. sequences $\theta = 1$). Such θ exists under general conditions (cf [4], Sec. 3.7)

Clusters of exceedances of high levels may be defined in various ways, the most obvious being as runs of consecutive exceedances. However if $\Delta(v_n)$ holds for a given sequence $\{v_n\}$ of levels the following "block definition" is more convenient (and often asymptotically equivalent to that using runs). Choose integers $k_n \longrightarrow \infty$, $k_n = o(n)$, satisfying

$$k_n(\alpha_{n,\ell_n} + \ell_n/n) \longrightarrow 0. \tag{3.2}$$

Write $r_n=[n/k_n]$ and divide the integers $(1...n)$ into consecutive blocks

$$J_i = \{(i-1)r_n+1,\ (i-1)r_n+2,\dots,ir_n\} \quad 1 \leq i \leq k_n$$

$$J_{k_n+1} = \{k_nr_n+1,\ k_nr_n+2,\dots,n\}$$

and regard the exceedances (if any) in a block as forming a cluster. The choice of k_n (or equivalently r_n) is flexible, subject to the growth restriction (3.2).

Obviously this block definition can count a run of consecutive exceedances as two or more clusters, if the run straddles more than one block so that the block definition is less natural in some cases. However it can be more appropriate than the runs definition for sequences with high local variability. In any case we use the block definition for its convenience. Further if the block J_i contains exceedances the first point, $(i-1)r_n+1$, of J_i will be regarded as the <u>location</u> of the cluster in J_i.

Define now a <u>point process</u> P_n <u>of normalized cluster positions</u>, i.e. consisting of the points $\{((i-1)r_n+1)/n,\ 1 \leq i \leq k_n: M(J_i) > u_n\}$ where $M(E)$ is written for $\max\{X_j: j\epsilon E\}$. Associate with each point $((i-1)r_n+1)/n$ of P_n the corresponding maximum excess $M(J_i)-u_n = \max\{(X_j-u_n)_+: j \in J_i\}$, giving the (again technically "marked") point process P_n^* of peak excess values above the level in the clusters. We show that a result like Theorem 2.1 holds in the dependent case with P_n^* replacing N_n^*. In fact it may be shown (cf [5]) that for i.i.d. sequences N_n and P_n are asymptotically equivalent in a strong sense as also are N_n^* and P_n^* so that P_n and P_n^* generalize N_n and N_n^*.

It was shown in Section 1 that in the iid case the exceedance point process N_n has a Poisson limit with intensity τ. This result is simply generalized (cf. [1]) to show that under dependence P_n has a Poisson limit (with intensity $\theta\tau$) whereas N_n itself has a Compound Poisson limit whose events occur at the positions of clusters (i.e. points of P_n), with multiplicities given by (limiting) cluster sizes. However the (limiting) distribution for multiplicities need not have a GP form and need not be totally determined by the (tail of the) marginal d.f. F of the X_i. On the other hand it will be shown below that P_n^* has a CP($\theta\tau$,G) distributional limit where G is obtained from the tail of F via (1.4). Thus G has GP form and the dependence influences the limit only through the factor θ in the intensity of the underlying Poisson Process.

The Compound Poisson limit for P_n^* will be obtained by considering the asymptotic behavior of the maximum in a cluster and showing that the clusters are essentially independent. The specific basic results needed are contained in the following lemma.

Lemma 3.1. Let $\{X_n\}$ be stationary with extremal index $\theta > 0$ and marginal d.f. F satisfying (1.4), and let $\Delta(v_n)$ hold with $v_n=u_n+xg(u_n)$, each $x \geq 0$, where u_n satisfies (2.1). Let $\{k_n\}$ satisfy (3.2) and write $a_n = (g(u_n))^{-1}$. Then (with $r_n = [n/k_n]$), as $n \longrightarrow \infty$

(i) $P\{a_n(M_{r_n}-u_n) > x\} \sim \frac{\theta\tau}{k_n}\overline{G}(x)$, $x \geq 0$

(ii) $P\{a_n(M_{r_n}-u_n) \leq x | M_{r_n} > u_n\} \longrightarrow G(x)$

(iii) $[\mathscr{E}\exp\{-sa_n(M_{r_n}-u_n)_+\}]^{k_n} \longrightarrow \exp\{-\theta\tau(1-\phi(s))\}$

where $\phi(s) = \int_0^\infty e^{-sx}dG(x)$

Proof: Since $n[1-F(u_n+a_n^{-1}x)] \sim \tau\overline{F}(u_n+xg(u_n))/\overline{F}(u_n) \longrightarrow \tau\overline{G}(x)$ and $\{X_n\}$ has extremal index θ, it follows that

$$P\{M_n \leq u_n+a_n^{-1}x\} \longrightarrow \exp(-\theta\tau\overline{G}(x)).$$

But it follows in a standard way from the mixing condition $\Delta(v_n)$ (cf. [1, Lemma 2.3]) that $P\{M_n \leq v_n\} - P^{k_n}\{M_{r_n} \leq v_n\} \longrightarrow 0$ so that $P^{k_n}\{M_{r_n} \leq u_n+a_n^{-1}x\} \longrightarrow \exp(-\theta\tau\overline{G}(x))$ from which (i) readily follows. The left hand side of (ii) is

$$1 - P\{M_{r_n} > u_n+a_n^{-1}x\}/P\{M_{r_n} > u_n\} = 1 - \frac{\theta\tau}{k_n}\overline{G}(x)/\frac{\theta\tau}{k_n}(1+o(1))$$

by (i), so that (ii) follows. Finally (iii) follows by the first calculation of Theorem 2.1 with M_{r_n} replacing X_i, using (ii). □

The main result now follows in a similar way to Theorem 2.1 on using approximate independence between the clusters.

<u>Theorem 3.2</u>. Let $\{X_n\}$ be stationary with extremal index $\theta > 0$, and marginal d.f F satisfying (1.4), and let $\Delta(v_n)$ hold with $v_n = u_n + xg(u_n)$ each $x \geq 0$, where u_n satisfies (2.1). Let $k_n \longrightarrow \infty$ satisfy (3.2), $r_n = [n/k_n]$ and $a_n = (g(u_n))^{-1}$. Then the point process P_n^* of peak values above u_n satisfies $a_nP_n^* \xrightarrow{d} P^*$ where P^* is $CP(\theta\tau, G)$ and G is as in (1.4).

Proof: Let I be a subinterval of (0,1] and $J_i = \{(i-1)r_n+1,\ldots,ir_n\}$ as defined above, $1 \le i \le k_n$. Then it follows from the mixing conditions along the same lines as Lemma 2.2 of [3] (or Lemma 2.2 of [1]) that

$$\mathcal{E}\exp\{-sa_nP_n^*(I)\} - \prod_{\{i:J_i\subset I\}} \mathcal{E}\exp\{-sa_nP_n^*(J_i)\} \longrightarrow 0$$

as $n \longrightarrow \infty$, so that

$$\mathcal{E}\exp\{-sa_nP_n^*(I)\} = [\mathcal{E}\exp\{-sa_n(M_{r_n}-u_n)_+\}]^{k_nm(I)(1+o(1))}$$

$$\longrightarrow \exp\{-\theta\tau m(I)(1-\phi(s))\}$$

by Lemma 3.1, where $\phi(s) = \int_0^\infty e^{-sx}dG(x)$. Hence $a_nP_n^*(I) \xrightarrow{d} P^*(I)$ where P^* is $CP(\theta\tau,G)$ on (0,1]. Now if $I_1,\ldots,I_k$ are disjoint subintervals of (0,1] it follows also as in Lemma 2.2 of [3] (or Lemma 2.2 of [1]) that

$$\mathcal{E}\exp\{-a_n \sum_{j=1}^{k} s_jP_n^*(I_j)\} - \prod_{j=1}^{k} \mathcal{E}\exp\{-a_nP_n^*(I_j)\} \longrightarrow 0$$

from which it follows that

$$(a_nP_n^*(I_1)\ldots a_nP_n^*(I_k)) \xrightarrow{d} (P^*(I_1)\ldots P^*(I_k))$$

and hence that $a_nP_n^* \xrightarrow{d} P^*$. □

Finally we reiterate that this result is one of many which can be obtained involving different aspects of cluster structure. For example the Compound Poisson limit for N_n was cited above, the multiplicities corresponding to cluster sizes. More complicated functions - such as the sum of powers of excess values in a cluster - may also be considered and will lead to Compound Poisson limits. Such cases may be useful in applications where damage from high levels (e.g. high pollution episodes) may be modeled as a specific function of the excess values. However the case of (excess) peak values in a cluster is especially important (e.g. in describing severe floods, where damage can well be a function of flood level). Theorem 3.2 establishing the POT approximation is particularly useful since the multiplicity distribution then depends only on the marginal d.f. F and moreover is known to have GP form.

References

[1] Hsing, T., Hüsler, J., Leadbetter, M.R., "On the exceedance point process for a stationary sequence", Prob. Theor. Rel. Fields, 78, 97-112, 1988.

[2] Kallenberg, O., Random Measures, Academic Press, N.Y. 1976

[3] Leadbetter, M.R. and Hsing, T., "Limit theorems for strongly mixing stationary random measures" To appear in J. Stoch. Proc. and Applns., 1990.

[4] Leadbetter, M.R., Lindgren, G., Rootzén, H., "Extremes and related properties of random sequences and processes", Springer Statistics Series, 1983.

[5] Leadbetter, M.R. and Nandagopalan, S., "On exceedance point processes under mild oscillation restrictions" Extreme Value Theory, J. Hüsler and R.-D. Reiss (eds.) Proc. Oberwolfach 1987, Springer Lecture Notes in Statistics, No. 51, 1988.

[6] Pickands, J., "Statistical inference using extreme order statistics" Ann. Statist. 3, 1975, p. 119-131.

[7] Smith, R.L., "Threshold methods for sample extremes" in Statistical Extremes and Applications, J. Tiago de Oliveira Ed., NATO ASI Series, Reidel, Dordrecht, 1985, p. 623-638.

[8] Smith, R.L. "Estimating tails of probability distributions" Ann. Statist. 15, 1987, p. 1174-1207.

256. E. Mayer-Wolf, A central limit theorem in nonlinear filtering, Apr. 89.

257. C. Houdré, Factorization algorithms and non-stationary Wiener filtering, Apr. 89, Stochastics, to appear.

258. C. Houdré, Linear Fourier and stochastic analysis, Apr. 89.

259. G. Kallianpur, A line grid method in areal sampling and its connection with some early work of H. Robbins, Apr. 89. *Amer. J. Math. Manag. Sci.*, 1989, to appear.

260. G. Kallianpur, A.G. Miamee and H. Niemi, On the prediction theory of two-parameter stationary random fields, Apr. 89. *J. Multivariate Anal.*, 32, 1990, 120-149.

261. I. Herbst and L. Pitt, Diffusion equation techniques in stochastic monotonicity and positive correlations, Apr. 89.

262. R. Selukar, On estimation of Hilbert space valued parameters, Apr. 89. (*Dissertation*)

263. E. Mayer-Wolf, The noncontinuity of the inverse Radon transform with an application to probability laws, Apr. 89.

264. D. Monrad and W. Philipp, Approximation theorems for weakly dependent random vectors and Hilbert space valued martingales, Apr. 89.

265. K. Benhenni and S. Cambanis, Sampling designs for estimating integrals of stochastic processes, Apr. 89.

266. S. Evans, Association and random measures, May 89.

267. H.L. Hurd, Correlation theory of almost periodically correlated processes, June 89.

268. O. Kallenberg, Random time change and an integral representation for marked stopping times, June 89. *Probab. Th. Rel. Fields*, accepted.

269. O. Kallenberg, Some uses of point processes in multiple stochastic integration, Aug. 89.

270. W. Wu and S. Cambanis, Conditional variance of symmetric stable variables, Sept. 89.

271. J. Mijnheer, U-statistics and double stable integrals, Sept. 89.

272. O. Kallenberg, On an independence criterion for multiple Wiener integrals, Sept. 89.

273. G. Kallianpur, Infinite dimensional stochastic differential equations with applications, Sept. 89.

274. G.W. Johnson and G. Kallianpur, Homogeneous chaos, p-forms, scaling and the Feynman integral, Sept. 89.

275. T. Hida, A white noise theory of infinite dimensional calculus, Oct. 89.

276. K. Benhenni, Sample designs for estimating integrals of stochastic processes, Oct. 89. (*Dissertation*)

277. I. Rychlik, The two-barrier problem for continuously differentiable processes, Oct. 89.

278. G. Kallianpur and R. Selukar, Estimation of Hilbert space valued parameters by the method of sieves, Oct. 89.

279. G. Kallianpur and R. Selukar, Parameter estimation in linear filtering, Oct. 89.

280. P. Bloomfield and H.L. Hurd, Periodic correlation in stratospheric ozone time series, Oct. 89.

281. J.M. Anderson, J. Horowitz and L.D. Pitt, On the existence of local times: a geometric study, Jan. 90.

282. G. Lindgren and I. Rychlik, Slepian models and regression approximations in crossing and extreme value theory, Jan. 90.

283. H.L. Koul, M-estimators in linear models with long range dependent errors, Feb. 90.

284. H.L. Hurd, Almost periodically unitary stochastic processes, Feb. 90.

285. M.R. Leadbetter, On a basis for 'Peaks over Threshold' modeling, Mar. 90.

286. S. Cambanis and E. Masry, Trapezoidal stratified Monte Carlo integration, Mar. 90.

287. M. Marques and S. Cambanis, Dichotomies for certain product measures and stable processes, Mar. 90.

288. M. Maejima and Y. Morita, Trimmed sums of mixing triangular arrays with stationary rows, Mar. 90.

289. S. Cambanis and M. Maejima, Characterizations of one-sided linear fractional Lévy motions, Mar. 90.

290. N. Kôno and M. Maejima, Hölder continuity of sample paths of some self-similar stable processes, Mar. 90.